

A classification of DEA models when the internal structure of the Decision Making Units is considered

Lorenzo Castelli · Raffaele Pesenti · Walter Ukovich

© .Springer Science+Business Media, LLC 2008

Abstract We classify the contributions of DEA literature assessing Decision Making Units (DMUs) whose internal structure is known. Starting from an elementary framework, we define the main research areas as shared flow, multilevel and network models, depending on the assumptions they are subject to. For each model category, the principal mathematical formulations are introduced along with their main variants, extensions and applications. We also discuss the results of aggregating efficiency measures and of considering DMUs as submitted to a central authority that imposes constraints or targets on them. A common feature among the several models is that the efficiency evaluation of the DMU depends on the efficiency values of its subunits thereby increasing the discrimination power of DEA methodology with respect to the black box approach.

Keywords Efficiency evaluation · Data envelopment analysis · Networks · Hierarchy · Multi-stage production processes

Data Envelopment Analysis (DEA) has been a standard tool for evaluating the relative efficiencies of Decision Making Units (DMUs) since the paper of Charnes et al. (1978) based on the seminal work of Farrell (1957). Some underlying assumptions are common to classical DEA models. The *efficiency* of a DMU is defined as the weighted ratio of the outputs (products or outcomes) yielded by the DMU over the inputs (resources used or consumed). All DMUs considered are *homogeneous*, i.e., they all have the same types of inputs and outputs, and are *independent*, i.e., no constraint binds input and output levels of a DMU with the inputs and outputs of other DMUs. Furthermore, DMUs are seen as *black boxes*, i.e., their internal structures are not considered. As a consequence, generally, there is no clear

L. Castelli (✉) · W. Ukovich
Dipartimento di Elettrotecnica, Elettronica e Informatica, Università degli Studi di Trieste,
Via A. Valerio 10, 34127 Trieste, Italy
e-mail: castelli@units.it

R. Pesenti
Dipartimento di Matematica Applicata, Università Ca' Foscari di Venezia, Dorsoduro 3825/E, 30123
Venice, Italy

evidence of the transformations to which the inputs are subject to within the considered units.

In the last two decades, several authors have explored the possibility of measuring efficiency relative to the subprocesses of DMUs within the DEA framework. These authors abandon the black box perspective in the assumption that, in some particular contexts, the knowledge of the internal structure of DMUs can give further insights for the DMU performance evaluation. As an example, such knowledge allows to determine whether better performances can be theoretically obtained by merging the technologies of some substructures of the observed DMUs. In addition, assessing the efficiency of each of the subunits might prevent the inefficiency of some of them to be compensated by the efficiency of other subunits.

The aim of this paper is to survey the models that consider internal structures of DMUs. The main rationale of the classification is driven by the realization that three groups of models are different generalizations of the same elementary formulation. As it appears from the synoptic table in the [Appendix](#) (see [Table 1](#)), the references form nearly closed clusters for each of the identified model categories. In this work we describe the commonalities and differences between these models and show how they relate to the basic formulation.

In particular, we analyze a specific model by comparing a set of homogeneous and independent DMUs, each composed of a set of Decision Making SubUnits (DMSUs). Subunit, component and (sub)process are other synonyms of DMSU. Each subunit is allowed to perform a unique function or activity. Only to keep the notation simple, we also assume that all the DMUs under comparison have the same internal structure.

We define a DMU as *elementary* if its internal structure complies with the following assumptions:

Assumptions

1. All the subunits of the DMU do not have shared inputs and shared outputs, i.e., the DMU does not have the opportunity to decide how to allocate its inputs or outputs among its subunits in order to maximize its efficiency (Cook et al. 2000).
2. Any input (output) of the DMU is also an input (output) of one of its subunits.
3. No intermediate flows among DMSUs exist. In other words, the output of a DMSU cannot be the input of another DMSU (and also cannot re-enter the same DMSU).

In the black box approach it is customary that the flow levels entering and exiting a DMU are not problem decision variables, but problem data. However, a different allocation of the common flows among the subunits may lead to different DMU efficiency values when two or more DMU components use the same type of input and/or produce the same type of output. We refer to *shared flow models* when Assumption 1 no longer holds (see [Fig. 1](#)). This situation may occur when DMUs are divided into different components that require common resources (e.g., money) or produce goods or services obtained through the synergy and collaboration among them (e.g., the quality of service provided to customers). Assumption 1 implies that the components of an elementary DMU do not compete for the same resource and do not synergically yield the same product. It follows that the combined presence of Assumption 1 and Assumption 3 signifies that all the subunits of an elementary DMU are independent. But they are neither requested nor prevented to be homogeneous. In fact, Assumption 1 allows subunits to have inputs and outputs of different type. We refer to *multilevel models* when Assumption 2 is dropped, i.e., when DMU inputs (outputs) are not necessarily inputs (outputs) of its subunits (see [Fig. 2](#)). When Assumption 3 is neglected we

refer to *network* models. Here DMUs have at least one output of a DMSU which is an input of a different DMSU (see Fig. 3). In this case, the subunits are neither homogeneous nor independent. Besides modeling subprocesses, these models are of interest because they allow to describe systems where DMUs are organized in networks where the outputs of some of them become inputs for other DMUs. This framework may encompass manufacturing production systems, and in general supply chains, in which some DMUs yield intermediate products that feed other DMUs. The same approach also includes dynamic DEA models in which some outputs at period t are inputs in the next period, $t + 1$ (Färe and Grosskopf 2000).

A further line of research considers DMUs as components of a greater structure which is interested in maximizing its future efficiency by either re-allocating resources or fixing targets to its subunits. Finally, we discuss the problem of consistently aggregate efficiency indexes across subunits to obtain a comprehensive DMU measure of efficiency.

In the following sections, we first describe the formulation to maximize the relative efficiency of an elementary DMU (Sect. 1). Then we introduce the basic reference models (typically with constant returns to scale) for shared flow (Sect. 2), multilevel (Sect. 3) and network (Sect. 4) models. We provide interpretations and applications proposed by different authors, and specify the possible variations from the basic model. In Sect. 5, we present the main results on the aggregation of efficiency indexes and on the resource re-allocating and target setting models. Finally, in Sect. 6 conclusions are drawn. Throughout the paper we assume that the reader is familiar with at least the seminal works on DEA (see, e.g., Charnes et al. 1978; Banker et al. 1984), as we will not define or justify basic concepts such as, e.g., positive non-Archimedean value ε , slack variables, production set, virtual inputs and outputs, returns to scale, allocative and technical efficiencies.

1 Elementary model

For each elementary DMU k (i.e., whose internal structure follows Assumptions 1–3) let us define

- i, j, r : the indexes of the generic input, output, and DMSU, respectively,
- $X_k^r = \{x_{ik}^r\}$: the vector of the inputs of DMSU r ,
- $Y_k^r = \{y_{jk}^r\}$: the vector of the outputs of DMSU r ,
- $v^r = \{v_i^r\}$: the vector of weights of the inputs of DMSU r ,
- $\mu^r = \{\mu_j^r\}$: the vector of weights of the outputs of DMSU r .

For an elementary DMU 0 belonging to a set of N homogeneous and independent DMUs with the same internal structure, the input-oriented version of the envelopment problem with constant returns to scale can be written as follows:

$$\theta_0^* = \min \theta_0 - \varepsilon \left(\sum_r \left(\sum_i s_i^{r-} + \sum_j s_j^{r+} \right) \right) \quad (1a)$$

$$\sum_k \lambda_k^r x_{ik}^r = \theta_0 x_{i0}^r - s_i^{r-} \quad \forall i, r \quad (1b)$$

$$\sum_k \lambda_k^r y_{jk}^r = y_{j0}^r + s_j^{r+} \quad \forall j, r \quad (1c)$$

$$\lambda_k^r, s_i^{r-}, s_j^{r+} \geq 0 \quad \forall i, j, k, r \quad (1d)$$

where λ_k^r is the multiplier of DMSU r belonging to DMU k , and s_i^{r-}, s_j^{r+} are the slack variables. If the DMU internal structure is unknown, the multipliers would not be associated to DMSUs, but to DMUs only, i.e., $\lambda_k^r = \lambda_k \forall r, \forall k$ where λ_k is the multiplier of DMU k . The dual formulation of Model (1) is:

$$e_0^* = \max \sum_{j,r} \mu_j^r y_{j0}^r \quad (2a)$$

$$\sum_{i,r} v_i^r x_{i0}^r = 1 \quad (2b)$$

$$\sum_j \mu_j^r y_{jk}^r \leq \sum_i v_i^r x_{ik}^r \quad \forall k, r \quad (2c)$$

$$v_i^r, \mu_j^r \geq \varepsilon \quad \forall i, j, r. \quad (2d)$$

In Model (2) the maximum relative efficiency e_0^* is assessed by comparing DMU 0 with all the existing subunits. As shown in Yang et al. (2000), Castelli et al. (2004), and Kao (2008) e_0^* is equal to the maximum relative efficiency of its subunits, and DMU 0 is:

- weakly efficient if and only if there exists at least one of its subunits which is weakly efficient relative to the corresponding subunits of other DMUs;
- CCR-efficient if and only if each of its subunits is CCR-efficient relative to the corresponding subunits of other DMUs.

It is worth noting that Model (2) can also be directly derived from the formulation of the relative efficiency maximization of e_0 as a comparison of DMU 0 with all the existing DMUs (Castelli et al. 2004), as it is customary in the classical DEA framework.

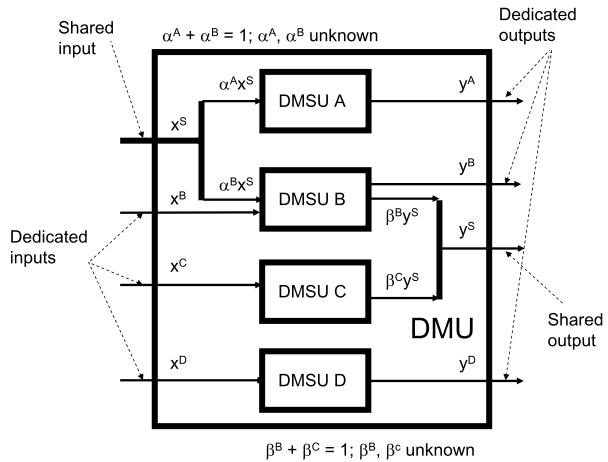
Also Färe and Primont (1984) proposed a multiple input and single output elementary configuration. Relying on the Farrell (1957) output-based efficiency measure, the authors constructed a reference technology for DMUs using their subunit data. Next, this efficient technology is compared against the reference frontier of the subunits, i.e., as if the subunits were independent DMUs and not part of a larger DMU. Kao (2000) generalized their models for cases of multiple outputs and multiple inputs.

2 Shared flow models

While the total amount of each flow entering (or exiting) the whole DMU is always known, as it is customary in DEA models, the amount of shared flow allocated to each component may be considered as a decision variable to maximize the DMU efficiencies (see Fig. 1).

Beasley (1995) introduced one of the first examples of a shared flow DEA model. Even if it was not originally referred to as such, it is nowadays acknowledged as part of the literature on the subject. It applies to departments of different universities devoted to the same disciplines. The considered departments are homogeneous and independent DMUs. Within each of them, the teaching and research activities clearly define two different separable functions. One of the DMU inputs, research income, is specifically *dedicated* to the research function. The other DMU inputs, general and equipment expenditure, are *shared (joined)* between the two functions. DMU outputs are split, i.e., no shared outputs exist: the number of undergraduates and of taught postgraduates are outputs of the teaching function; the number of research postgraduates, research income, and research rating are outputs of the research function. Other shared flow formulations are applied in several domains. By introducing variable returns to scale, Tsai and Molinero (2002) evaluate the performance of National

Fig. 1 A shared flow DMU: DMSUs A and B are not independent because they compete for the same shared resource. Similarly DMSUs B and C are not independent because of the shared output. DMSU D is independent of DMSUs A, B and C



Health Service (NHS) trusts in England, and Diez-Ticio and Mancebon (2002) assess the efficiency of Spanish Police Service. A case study on Canadian bank branches is presented by Cook et al. (2000), Cook and Hababou (2001) and Cook and Zhu (2005, Chap. 6) where different weights on shared inputs and an additive objective function are considered. Shared flow models with panel data and non-discretionary inputs are introduced by Jahanshahloo et al. (2004a, 2004b) and Amirteimoori and Nashtaei (2006) in a case study on Iranian commercial bank branches. Panel data are considered to measure possible progress and regress (see Tulkens and Eeckaut 1995) of the bank sector.

2.1 Formulation of shared flow models

Referring to r as the generic component of DMU k , now vectors X_k^r, Y_k^r, v^r , and μ^r introduced in Sect. 1 are defined as the vectors of dedicated inputs, dedicated outputs, weights of the dedicated inputs, and weights of dedicated outputs of component r , respectively. In addition, we define

- $X_k^S = \{x_{ik}^S\}$: the vector of shared inputs,
- $Y_k^S = \{y_{jk}^S\}$: the vector of shared outputs,
- $v^S = \{v_i^S\}$: the vector of weights of shared inputs,
- $\mu^S = \{\mu_j^S\}$: the vector of weights of shared outputs,
- $\alpha^r = \{\alpha_i^r\}$: the vector of proportions of the shared inputs allocated to component r ,
- $\beta^r = \{\beta_j^r\}$: the vector of proportions of the shared outputs attributed to component r .

With a little abuse of notation we also define $\alpha^r X_k^S$ as the column vector whose generic entry is $\alpha_i^r x_{ik}^S$. In this context, $\alpha_i^r x_{ik}^S$ is the amount of shared input i allocated to component r by DMU k to maximize its efficiency. When a shared input cannot be clearly divided among functions (e.g., general expenditure), then α_i^r can be seen as the proportion of the (virtual) value of the input i allotted to component r . Similarly, we define $\beta^r Y_k^S$ as the column vector whose generic entry is $\beta_j^r y_{jk}^S$ where β_j^r is always seen as the proportion of the (virtual) value of output j that can be attributed to component r because it is assumed that no component can produce a shared output by itself but needs synergy with other components. As an example, the quality of service level provided by an organization to its customers depends on the degree of collaboration and integration among its subdivisions, each of them sharing

with other subunits the responsibility for such output. When outputs common to different components are produced without the need of synergy among them, the literature refers to them as *overlapping* outputs (see Sect. 2.2.5 for details).

2.1.1 Primal formulation

Consider, for the sake of simplicity, the case when shared outputs are not present. The efficiency of DMU k is expressed as

$$e_k = \frac{\sum_r \mu^r Y_k^r}{\sum_r v^r X_k^r + \sum_r v^S (\alpha^r X_k^S)},$$

the *partial* efficiency of the single component r is defined as

$$e_k^r = \frac{\mu^r Y_k^r}{v^r X_k^r + v^S (\alpha^r X_k^S)},$$

and the *aggregate* efficiency $\hat{e}_k = \sum_r q_k^r e_k^r$ as the weighted combination of the partial efficiencies of its components, where the weight q_k^r of each component r is

$$q_k^r = \frac{v^r X_k^r + v^S (\alpha^r X_k^S)}{\sum_p v^p X_k^p + \sum_p v^S (\alpha^p X_k^S)}.$$

Hence q_k^r is the fraction of DMU k total weighted inputs that are consumed by component r : $\sum_r q_k^r = 1 \forall k$. Also Yang et al. (2000) introduced the concept of partial efficiency measures but they applied it on an elementary model (see Sect. 1). The general model proposed by Beasley (1995) is

$$e_0^* = \max e_0 \quad (3a)$$

$$e_k^r \leq 1 \quad \forall k, r \quad (3b)$$

$$\sum_r \alpha_i^r = 1 \quad \forall i \quad (3c)$$

$$v_i^r, v_i^S, \alpha_i^r, \mu_j^r \geq \varepsilon \quad \forall i, j, r. \quad (3d)$$

Condition (3b) imposes that the partial efficiency of each DMU component cannot exceed 1. Beasley (1995) proves that when each DMU is free to allocate the value of the shared inputs among its different components, the aggregate efficiency \hat{e}_k and the efficiency e_k are coincident when maximized.

As for the classical DEA formulations, Model (3) can be rewritten as follows

$$e_0^* = \max \sum_r \mu^r Y_0^r \quad (4a)$$

$$\sum_r v^r X_0^r + \sum_r v^S (\alpha^r X_0^S) = 1 \quad (4b)$$

$$\mu^r Y_k^r \leq v^r X_k^r + v^S (\alpha^r X_k^S) \quad \forall k, r \quad (4c)$$

$$\sum_r \alpha_i^r = 1 \quad \forall i \quad (4d)$$

$$v_i^r, v_i^S, \alpha_i^r, \mu_j^r \geq \varepsilon \quad \forall i, j, r. \quad (4e)$$

Model (4) is not linear because of inequalities (4b)–(4c). When no shared inputs exist Model (4) easily reduces to the elementary Model (2) as $X_k^S = 0 \forall k$. Hence the terms $\sum_r v^S(\alpha^r X_0^S)$ in (4b) and $v^S(\alpha^r X_k^S)$ in (4c), and constraint (4d) are no longer necessary.

2.1.2 Dual formulation

Molinero (1996) and Molinero and Tsai (1997), proposed an approach dual to (3). In addition, the authors included *shared outputs*, i.e., outputs yielded synergically by two or more components. Their output oriented model is

$$e_0^* = \max \sum_r q_0^r \theta_0^r + \varepsilon \left(\sum_i \left(s_i^{S-} + \sum_r s_i^{r-} \right) + \sum_j \left(s_j^{S+} + \sum_r s_j^{r+} \right) \right) \tag{5a}$$

$$\sum_k \lambda_k^r x_{ik}^r = x_{i0}^r - s_i^{r-} \quad \forall i, r \tag{5b}$$

$$\sum_k \sum_r \lambda_k^r (\alpha_i^r x_{ik}^S) = x_{i0}^S - s_i^{S-} \quad \forall i \tag{5c}$$

$$\sum_k \lambda_k^r y_{jk}^r = \theta_0^r y_{j0}^r + s_j^{r+} \quad \forall j, r \tag{5d}$$

$$\sum_k \sum_r \lambda_k^r (\beta_j^r y_{jk}^S) = \sum_r \theta_0^r (\beta_j^r y_{j0}^S) + s_j^{S+} \quad \forall j \tag{5e}$$

$$\sum_r \alpha_i^r = 1 \quad \forall i \tag{5f}$$

$$\sum_r \beta_j^r = 1 \quad \forall j \tag{5g}$$

$$\sum_r q_0^r = 1 \tag{5h}$$

$$\lambda_k^r, q_0^r, \alpha_i^r, \beta_j^r, s_i^{r-}, s_i^{S-}, s_j^{r+}, s_j^{S+} \geq 0 \quad \forall i, j, r, k, \tag{5i}$$

where q_0^r are positive weights representing the relative importance of each component r for DMU 0, and θ_0^r are measures of the inefficiencies of the components of DMU 0. Actually, θ_0^r are the reciprocals of the *distance functions* defined by Shephard (1970). Note that in the models proposed by Molinero (1996) and Molinero and Tsai (1997) the slack variables s_i^{r-} , s_i^{S-} , s_j^{r+} , s_j^{S+} are not present. Here they are imposed for coherence with the standard DEA dual models (see, e.g., Cooper et al. 2000).

When the values α_i^r , β_j^r , and q_0^r are not decision variables but are fixed, still satisfying conditions (5f), (5g) and (5i), the dual of Model (5) is

$$e_0^* = \min \sum_r v^r X_0^r + \sum_r v^S (\alpha^r X_0^S) \tag{6a}$$

$$\mu^r Y_0^r + \mu^S (\beta^r Y_0^S) = q_0^r \quad \forall r \tag{6b}$$

$$\mu^r Y_k^r + \mu^S (\beta^r Y_k^S) \leq v^r X_k^r + v^S (\alpha^r X_k^S) \quad \forall k, r \tag{6c}$$

$$v_i^r, v_i^S, \mu_j^r, \mu_j^S \geq \varepsilon \quad \forall i, j, r. \tag{6d}$$

The above model parallels the output oriented version of Model (4) when shared outputs are considered. Besides Model (6) being linear, the main difference between the two models is the presence of the multiple constraints (6b) instead of the single one $\sum_r \mu^r Y_0^r + \sum_r \mu^S (\beta^r Y_0^S) = 1$. This latter constraint is a relaxation of the former ones because $\sum_r q_0^r = 1$. Conditions (6b) state a precise relationship between the relative importance attributed to a component and the optimal amount of outputs allocated to it (respectively, the optimal amount of allocated inputs if an input oriented model is considered). Then, conditions (6b) justify the choice in Beasley (1995) of expressing the weight q_k^r of the component r in the aggregated efficiency as equal to the fraction of DMU k total weighted inputs that are consumed by component r . Without conditions (6b), such a choice might appear arbitrary, although reasonable.

2.2 Extensions

Many authors have extended Models (4) and (5). Common features of the different variants are that the aggregate efficiency of a DMU cannot exceed unity, and that a DMU is efficient if and only if it is efficient in all its components. In this section, we describe the peculiarity of each available modeling advance.

2.2.1 Weight restrictions

Beasley (1995) himself does not present Model (3), but he incorporates the additional constraints

$$(v^S, v^r, \forall r) \in \Omega_{in} \quad (7)$$

$$(\mu^r, \forall r) \in \Omega_{out} \quad (8)$$

where the sets Ω_{in} and Ω_{out} are *assurance regions* as defined in Thompson et al. (1990). Constraints (7) and (8) involve value judgements concerning the proportions α^r and the weights μ^r , and v^r of the different DMU components. They are not strictly necessary for the definition of a shared flow DEA model, but might prevent the model from yielding unreasonable results. In this context, Beasley (1995) provides an example where, in the absence of constraints (7)–(8), one research postgraduate was worth about 880000 undergraduates for a given department.

2.2.2 Variable returns to scale

Molinero and Tsai (1997) proved that the feasible solutions of (5) define a convex set and (5a) is a convex function. Tsai and Molinero (2002), considering university departments as a reference example, introduced and discussed a variable returns to scale version of (5). The efficiency of each component r of DMU k is then defined as

$$e_k^r = \frac{\mu^r Y_k^r + \mu^S (\beta^r Y_k^S)}{v^r X_k^r + v^S (\alpha^r X_k^S) + \delta_k^r} \quad (9)$$

where the variable δ_k^r is unrestricted and its optimal value defines the component's returns to scale status. The aggregate efficiency of DMU k is

$$e^r = \frac{\sum_r \mu^r Y_k^r + \sum_r \mu^S (\beta^r Y_k^S)}{\sum_p v^p X_k^p + \sum_p v^S (\alpha^p X_k^S) + \sum_p \delta_k^p}. \quad (10)$$

Note that the optimal value of $\sum_p \delta_k^p$ may be zero even if some or all elements in the sum are different from zero. In this case, DMU k may appear to be operating under constant returns to scale and technically efficient when analyzed as a black box but, when its individual components are analyzed, it may be found scale inefficient in each of its activities (Tsai and Molinero 2002). It follows that a DMU, that is efficient when considered as a black box, may be inefficient when its different components are taken into account, independently of its returns to scale status.

2.2.3 Different weights on shared inputs

Cook et al. (2000) allow a same shared input i to be weighted differently by the subunits of the same DMU. The rationale behind such a choice is that different components may disagree on the importance of a same input. Consequently, the shared flow model as in Cook et al. (2000) includes in (4b)–(4c) a set of vectors v^{Sr} , one for each component r , instead of a single one. Also, a change of variables is proposed. In particular, let $i = 1, \dots, s$ be the index of the shared inputs, then $\bar{v}_i^{Sr} = v_i^{Sr} \alpha_i^r$ for $i = 1, \dots, s - 1$ and $\bar{v}_s^{Sr} = v_s^{Sr} (1 - \sum_{i=1}^{s-1} \alpha_i^r)$. Because of these new variables, the authors obtain a linear model. The terms $v_r^S (\alpha^r X_k^S)$ in conditions (4b)–(4c) become $\bar{v}^{Sr} X_k^S$ and $v_r^S \geq \varepsilon$ in (4e) turns $\bar{v}_i^{Sr} \geq \varepsilon \alpha_i^r$. Unfortunately, non-linearity may arise again when additional constraints concerning value judgements as (7)–(8) are necessary. If such judgements are expressed also in terms of v_r^S , the variable substitution may not lead to a linear model.

2.2.4 Additive objective function

Cook and Hababou (2001) present variables and constraints as in Cook et al. (2000) but differ in the objective function. They formulate an additive objective function representing an aggregate measure of the efficiencies of all the DMU subunits. In the classical additive DEA models (Charnes et al. 1985), a possible measure of the inefficiency of DMU k is given by the difference between the weighted sum of the inputs minus the weighted outputs of DMU k . Here Cook and Hababou (2001) suggested a multiobjective approach where the partial inefficiencies of all components are considered. For each subunit, the weighted sum of its inputs minus the weighted sum of its outputs is considered. In particular, the authors minimize the maximum partial inefficiency in order to give equal importance to each component, i.e., their objective function is

$$\min \max \{v^r X_k^r + v^{Sr} (\alpha^r X_k^S) - \mu^r Y_k^r : \forall r \text{ subunit of DMU}_0\}. \quad (11)$$

Finally, the authors linearize their model with the same variable changes proposed in Cook et al. (2000).

2.2.5 Overlapping outputs

Cook and Green (2004) deal with a manufacturing multi-plant company and point out that some outputs of different components of the same DMU can partially overlap, i.e., some outputs may be common to different components. In particular, each DMSU can yield a given amount of overlapping output j , with no need of synergy with the other components. Hence there is no possibility of attributing the considered amount to the other subunits. From this point of view, the overlapping outputs are different from the shared outputs considered in Molinero and Tsai (1997); Molinero (1996), and Jahanshahloo et al. (2004b). In fact,

Cook and Green (2004) cannot approach what they call the *overlap problem* by introducing variables β^r as in (5) to determine which proportions of shared outputs are attributed to each component: the efficiency of a single subunit r remains $e_k^r = \frac{\mu^r Y_k^r}{\nu^r X_k^r + \nu^{Sr} (\alpha^r X_k^S)}$ and, consequently, the aggregate efficiency of a whole DMU k is $e_k = \frac{\sum_r \mu^r Y_k^r}{\sum_r \nu^r X_k^r + \sum_r \nu^{Sr} (\alpha^r X_k^S)}$. However, the shared inputs are no longer allocated to the components because such task could hardly be performed without introducing some ambiguities due to the component overlapping. Shared inputs are allocated directly to the outputs. In particular, consider Model (3) and the extension proposed in Cook et al. (2000). Cook and Green (2004) introduced a new set of variables α_i^j as the proportions of the shared inputs i allocated for outputs j . In addition, they replaced condition (3c) with $\sum_j \alpha_i^j = 1$, for all i . Finally, they defined α_i^r as $\alpha_i^r = \sum_{j \in O^r} \alpha_i^j$, where O^r is the set outputs of subunit r . Note that now, in general, $\sum_r \alpha_i^r \geq 1$.

The allocation of shared inputs directly to outputs was originally introduced in Färe et al. (1997). Even though the concept of DMSU is not explicitly mentioned, it can easily be inferred since one input can be allocated among various outputs.

2.2.6 Core business identification

Cook and Green (2004) and Cook and Zhu (2005, Chap. 11) addressed the problem of determining in which areas a DMU would perform better. Such areas form the *core business* of a DMU and should be privileged even at the cost of possibly forcing the DMU to abandon the components with less satisfactory performances. To this aim, Cook and Green (2004) modified the objective function of Model (3) and added assignment constraints (each DMU must have at least one component assigned and each component must be assigned to at least one DMU).

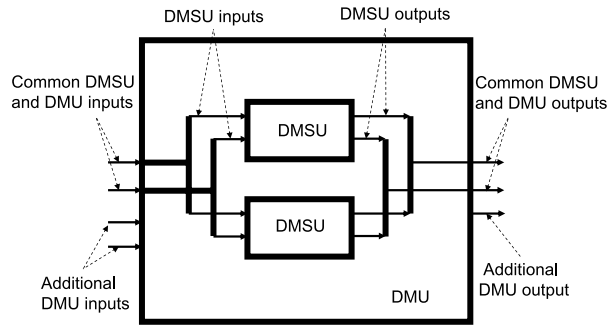
2.2.7 Panel data and non-discretionary inputs

Jahanshahloo et al. (2004b) extended the model proposed in Cook et al. (2000). They introduced shared outputs, considered panel data, and proved that the aggregate efficiency of the whole DMU is a convex combination of the efficiency of the DMU components even in the presence of shared outputs. In Jahanshahloo et al. (2004a), the same authors also introduced non-discretionary inputs that are not under decision maker control, but are considered as negative terms in the numerator of the fraction that describes a DMU efficiency value. They linearized the models proposed in Jahanshahloo et al. (2004b) by changing variables as in Cook et al. (2000).

3 Multilevel models

A different research area includes DMUs exhibiting autonomous activities that cannot be associated to any of its subunits. In other words, a DMU presents additional inputs/outputs not considered by its subunits. For example, in Cook et al. (1998), DMSUs are highway maintenance patrols and DMUs are the districts in which the maintenance patrols are grouped. The subunits have traffic and road conditions as possible inputs, while DMUs may include additional inputs that can be applied only to districts such as the extent of privatization and district engineers' experience. The same authors also introduced possible applications of their model to power plants and hospitals. These models are defined as *multilevel models* (Cook et al. 1998) where the top level, referred to as level n DMU, includes independent

Fig. 2 A multilevel DMU: the DMU includes two homogeneous and independent subunits



and homogeneous subunits, referred to as level $n - 1$ DMUs. Recursively, the level $n - 1$ DMUs include smaller independent and homogeneous subunits, level $n - 2$ DMUs, and so on. Unlike shared flow models, the amount of input and output of each subunit is fixed. In this work, we introduce only two-level structures, and we simply refer to DMU for the level 2 DMU and to DMSU or subunit for level 1 DMUs (see Fig. 2).

By denoting $i; j; k$ as the indexes of the generic input, output, and DMU, respectively, the following notation is introduced:

- $R_k = \{r_k\}$: the set of indexes r_k of all DMSUs belonging to DMU k ,
- $X_k^r = \{x_{ik}^r\}$: the vector of the inputs of DMSU r_k ,
- $X_k = \{x_{ik}\}$: the vector of the additional inputs of DMU k ,
- $Y_k^r = \{y_{jk}^r\}$: the vector of the outputs of DMSU r_k ,
- $Y_k = \{y_{jk}\}$: the vector of the additional outputs of DMU k ,
- $v^1 = \{v_i^1\}$: the vector of weights of the inputs common to both DMSUs and DMUs,
- $v^2 = \{v_i^2\}$: the vector of weights of the additional inputs of DMUs,
- $\mu^1 = \{\mu_j^1\}$: the vector of weights of the outputs common to both DMSUs and DMUs,
- $\mu^2 = \{\mu_j^2\}$: the vector of weights of the additional outputs of DMUs.

Accordingly, the efficiency of a DMSU r_k is expressed as

$$e_k^r = \frac{\mu^1 Y_k^r}{v^1 X_k^r} \tag{12}$$

and the efficiency of a DMU k as

$$e_k = \frac{\mu^1 \sum_{r_k \in R_k} Y_k^r + \mu^2 Y_k}{v^1 \sum_{r_k \in R_k} X_k^r + v^2 X_k} \tag{13}$$

Cook et al. (1998) presented a unifying model for multilevel structures that assesses the efficiency of DMUs of different levels. The authors argue that the efficiency of a DMSU r_k should be evaluated only relative to those other subunits operating under the same conditions, in practice belonging to the same DMU k .

On the other hand, they also assert that the subunits in R_k should be taken into account when evaluating the efficiency of a DMU k . On the basis of these assumptions, Cook et al. (1998) proposed that the efficiency of a DMSU 0_0 in DMU 0 is evaluated through the following model

$$e_{0_0}^* = \max \mu^1 Y_{0_0}^0 \tag{14a}$$

$$v^1 X_{0_0}^0 = 1 \tag{14b}$$

$$\mu^1 Y_0^r \leq v^1 X_0^r \quad \forall r_0 \in R_0 \quad (14c)$$

$$v_i^1, \mu_j^1 \geq \varepsilon \quad \forall i, j. \quad (14d)$$

This is a classical DEA model that evaluates DMSU 0_0 relative only to subunits included in the same DMU 0 . The efficiency of a DMU 0 is evaluated through the following model:

$$e_0^* = \max \mu^1 \sum_{r \in R_0} Y_0^r + \mu^2 Y_0 \quad (15a)$$

$$v^1 \sum_{r \in R_0} X_0^r + v^2 X_0 = 1 \quad (15b)$$

$$\mu^1 \sum_{r \in R_k} Y_k^r + \mu^2 Y_k \leq v^1 \sum_{r \in R_k} X_k^r + v^2 X_k \quad \forall k \quad (15c)$$

$$\mu^1 Y_k^r \leq v^1 X_k^r \quad \forall k, \forall r \in R_k \quad (15d)$$

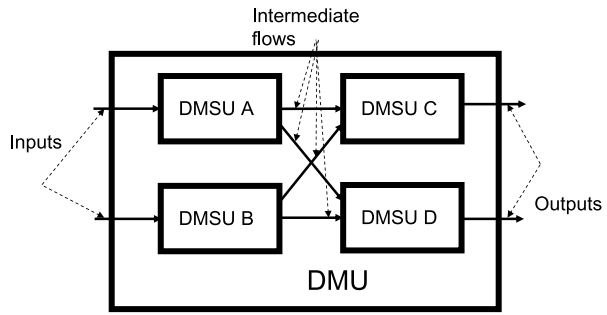
$$v_i^2, v_i^1, \mu_j^2, \mu_j^1 \geq \varepsilon \quad \forall i, j. \quad (15e)$$

This model compares DMU 0 with all other DMUs. It is different from the linear programming model considering DMUs as black boxes due to the presence of constraints (15d). These constraints take into account the DMU internal structure by imposing that their efficiency is related to the efficiencies of their subunits. In particular, constraints (15d) force that the optimal values for weights v_i^1, μ_j^1 are feasible for the DMSUs, i.e., the efficiency of each subunit should not exceed unity. Cook et al. (1998) presented a unifying model for multilevel structures that includes both (14) and (15). When the DMUs do not have additional inputs/outputs, Model (15) reduces to the elementary model (2). In such case, constraints (15c) turn out to be redundant since they are implied by constraints (15d). Cook and Green (2005) applied the hierarchical model described in Cook et al. (1998) to the evaluation of power plants.

3.1 Comparing subunits belonging to different DMUs

In Model (14) any subunit is compared only against the other subunits belonging to the same DMU. The rationale is that inputs received and decisions taken by each DMU influence the efficiency of its subunits, then comparing subunits belonging to different DMUs would be questionable. In fact, DEA models assess the efficiency of a DMU as a function of its distance from the production frontier defined by the other observed DMUs. In a mathematical programming perspective, DEA models determine the efficiency of a DMU with respect to the other DMUs. In an econometric perspective, the observed DMUs are a sample of a larger population, and DEA is a biased estimator of the efficiency of a DMU with respect to the unknown real production set (Simar and Wilson 2000). In both situations, the larger the sample is, the more likely the DMU under assessment is inefficient. Also, the average efficiency of the DMUs of the sample decreases (Zhang and Bartles 1998). This is why Staat (2001) invites to interpret very carefully possible differences in the efficiencies of subunits belonging to different DMUs when the cardinalities of sets R_k vary. Cook et al. (1998) proposed a way of correcting the possible biases by adjusting the efficiency of subunit r_k taking into account the size of the DMU k , the average efficiency of all the subunits in R_k , and the efficiency of DMU k . However, Staat (2002) points out that such a procedure returns different corrections for samples of equal size. He suggests to use bootstrap techniques (see, e.g., Simar and Wilson 2000) to overcome such deficiencies.

Fig. 3 Network DMU



4 Network models

In this section we describe models including intermediate flows between the subunits. The subunits are no longer independent, but are interdependent in the sense that part of the output produced by each of them may be used as an input by other ones. Interdependency among subunits leads to their non-homogeneity.

The basic model has been introduced by Färe and Whittaker (1995) and Färe and Grosskopf (1996) where each DMU is composed of two subunits and there is one intermediate flow: the output from the first subunit is used in the second one. Färe and Grosskopf (2000) extend the network considering more subprocesses (see also Färe et al. 2007). We propose a general formulation under the assumption that all DMUs have exactly the same internal structure: Model (16) assesses the relative efficiency of DMU 0 where r is the generic subunit, x_{i0}^r is the amount of the i -th input of DMU 0 entering subunit r , y_{j0}^r is the amount of the j -th output of DMU 0 produced by subunit r , and z_{l0}^r is the l -th intermediate flow of DMU 0: it is an output of subunit r used by subunit t (Fig. 3).

$$\theta_0^* = \min \theta_0 - \varepsilon \sum_r \left(\sum_i s_i^{r-} + \sum_l s_l^{r-} + \sum_j s_j^{r+} \right) \tag{16a}$$

$$\sum_k \lambda_k^r x_{ik}^r = \theta_0 x_{i0}^r - s_i^{r-} \quad \forall i, r \tag{16b}$$

$$\sum_k \lambda_k^r \sum_{t \in pred(r)} z_{lk}^t = \sum_{t \in pred(r)} z_{l0}^t - s_l^{r-} \quad \forall l, r \tag{16c}$$

$$\sum_k \lambda_k^r \left(y_{jk}^r + \sum_{t \in succ(r)} z_{jk}^t \right) = y_{j0}^r + \sum_{t \in succ(r)} z_{j0}^t + s_j^{r+} \quad \forall j, r \tag{16d}$$

$$\lambda_k^r \geq 0 \quad \forall k, r, \tag{16e}$$

where k is the generic DMU, and $pred(r)$ represents the set of predecessors of subunit r , i.e., the set of subunits which have at least one output used by subunit r . Similarly, $succ(r)$ is the set of successors of subunit r . For subunit r , input constraints are (16b) and (16c), output constraints are (16d). Expression (16b) and the y component of expression (16d) are also the input and output constraints of DMU 0, respectively. By introducing dummy subunits, Kao (2007) transformed a generic network of DMSUs into a series system where each stage in the series is of a parallel structure.

As pointed out in Färe and Grosskopf (1996), DMU 0 exhibits constant returns to scale and strong disposability of inputs and outputs since each subunit does the same.

Färe and Whittaker (1995) applied the model including intermediate flows to a dairy production problem and made a comparison with a classical DEA model, i.e., without intermediate production. The former model turned out to have greater discrimination power: only 23% of the DMUs are on the frontier compared to 65% when intermediate flows are not explicitly taken into account.

As a further application, Löthgren and Tambour (1999) estimated efficiency and productivity for a sample of Swedish pharmacies taking also into account customer satisfaction. The pharmacy technology is represented by a production and a consumption node. The production node yields (final) outputs (e.g., outpatient prescriptions) and also produces nonmarketable characteristics and attributes (e.g., the service level) that are considered as intermediate inputs of the consumption node. Together with external inputs (e.g., customer-service labor hours) the consumption node provides customer quality assessments on pharmacy service. Also Prieto and Zofio (2007) employ network DEA models as in Färe and Grosskopf (2000). They assess the economies of a set of countries belonging to the Organisation for Economic Co-operation and Development (OECD) with the aim of identifying best practices. Each national economy is described in terms of a network where different nodes use primary inputs to produce intermediate input and outputs, and satisfy final demand. Each node represents a basic economic sector, such as agriculture, manufacturing, construction, and services.

4.1 Extensions and special cases

A few authors have introduced variants of Model (16). They extend it to time-dependent processes and infer properties of the DMU efficiency when the internal structure has some special features. In addition, they address the efficiency evaluation of the subunits and link it with the whole DMU assessment.

4.1.1 Dynamic networks

Färe and Grosskopf (2000) considered the same production process in two successive periods with period-specific inputs and outputs. In addition, some of the outputs in the first period are used as inputs in the second. By representing the production processes as nodes or subunits, these time-intermediate products are the intermediate flows of a (dynamic) network whose relative efficiency may be evaluated using Model (16). Afterward, Chen (2008) focused on the temporal evolution of intermediate flows in a production network.

4.1.2 Multi-stage DMUs

Seiford and Zhu (1999) analyzed DMUs whose internal structure is composed of consecutive stages and considered each stage and the whole DMU as independent by evaluating their efficiencies using conventional DEA models. Conversely, Castelli et al. (2004) linked the DMU and subunit efficiencies of a two-stage process under constant returns to scale (see Fig. 3). Their main results may be summarized as follows:

- When only one stage exists, the internal structure reduces to the elementary one, thus the models of Sect. 1 apply.
- When the flow entering each subunit in the second stage is equal to the sum of the flows leaving the first stage, the maximum relative efficiency of the DMU under evaluation is assessed by comparing it with all the existing subunits, i.e., with each subunit belonging to each observed DMU.

- When the input virtual weight of each subunit in the second stage is equal to the sum of the output virtual weights of the subunits of the first stage, the maximum relative efficiency of the DMU under evaluation is equal to the product of the maximum relative efficiency of each single stage calculated according to Model (2).

Similar results were also provided by Amirteimoori and Shafiei (2006) and Kao and Hwang (2008). Furthermore, Chen and Zhu (2004) proposed a DEA framework which considers a two-stage process as efficient when each stage is efficient. Chen et al. (2007) proved the equivalence between the CCR version of the Chen and Zhu (2004) model and the Kao and Hwang (2008) model. Chen et al. (2008) showed that the analysis of subunits in series in the case of variable returns to scale cannot rely on radial measures of efficiency but needs an additive formulation.

In the assessment of multi-stage systems, Golany et al. (2006) simultaneously evaluated the efficiency of the aggregate system and of each DMSU within it. The authors acknowledged that their model is a special case of the Färe and Grosskopf network framework (see, e.g., Färe and Grosskopf 2000), but they used different aggregate efficiency measures as possible objective functions. In particular, Golany et al. (2006) proposed measures that allow to identify a Pareto optimal point, as the one that achieves the largest equal reduction in the efficiency of the DMSUs. The assessment of two-stage processes has been studied also relying on a game theory approach. In particular, Liang et al. (2006) compared a leader-follower and a cooperative relationship between DMSUs of a supply chain. Finally, Chen et al. (2006) proposed a DEA game model and proved the existence of numerous Nash equilibria efficiency plans for the DMSUs.

4.1.3 Evaluating the efficiency of DMSUs

The knowledge of the internal structure of the observed DMUs allows to determine whether better performances could be obtained by merging the technologies of their substructures. In addition, assessing the efficiency of each subunit might prevent that the inefficiency of some of them may be compensated by the efficiency of others. Hence the need, when intermediate flows exist, to develop models aiming at evaluating subunit efficiencies and at studying the influence of such values to the efficiency of the DMU the subunits belong to.

In this context, Castelli et al. (2001) introduced a DEA-like model to compare non-homogenous and interdependent subunits belonging to the same DMU. A given subunit r may be evaluated according to three different sets: (a) all the subunits homogeneous to it, (b) all the subunits of the DMU, and (c) with respect to a given output, all the subunits yielding that output. In this last case, the rationale is that these subunits, although not necessarily homogeneous, have a certain degree of commonality because they can be considered as potential substitutes for each other, as far as the production of that output is concerned. Thus the interest in comparing them. As a possible limitation, Lewis and Sexton (2004) pointed out that this approach may lead to small reference sets. In addition, Castelli et al. (2001) linked the subunits' and DMU efficiencies by defining an efficiency value W obtained by maximizing the product of the efficiency of the subunit under evaluation and the efficiency of the DMU it belongs to. In this way, subunits not only maximize their own efficiency, but also positively contribute to the efficiency of the whole system they are part of: the authors proved that a subunit seeking to optimize its W efficiency behaves with a benevolent attitude, i.e., being equal to other conditions, it also maximizes a combination of the efficiencies of the other subunits. Finally, they showed that the whole DMU is efficient if and only if all its subunits are W efficient.

Also Sexton and Lewis (2003) and Lewis and Sexton (2004) explicitly computed the efficiencies of the subunits. Model (16) is easily adapted to this task by optimizing the efficiency of subunit r belonging to DMU 0 and by adequately rewriting constraints (16b)–(16d). Both input and output oriented formulations are presented. In a simple case of DMUs composed of two subunits S_1 and S_2 in series, DMU 0 is defined as efficient when its output values are equal to the output values produced in the case that S_2 is efficient and uses the intermediate product levels that would have arrived at it had S_1 been efficient. When the internal structure of each DMU can be modeled as an acyclic direct graph, where the nodes represent the subunits and one origin and one destination node are added, Lewis and Sexton (2004) stated that the existence of a path from origin to destination along with every subunit is efficient is a necessary (but not sufficient) condition for the whole DMU to be efficient. As a consequence, it is possible that when considering the internal structure all DMUs under evaluation are inefficient.

5 Other areas of research

In this section we present studies for the efficiency evaluation of DMUs that take into account their internal structures but are not strictly related to the elementary formulations (1) and (2) or their extensions described in the previous sections.

5.1 Aggregate efficiency

Farrell (1957) introduced the concept of assessing the aggregate efficiency value of a group of subunits belonging to the same DMU. In this context, Blackorby and Russell (1999) provided a major result showing that consistent aggregation is feasible only under very strong restrictions. In particular, the index for the aggregate efficiency is defined as consistent when it can be derived as a continuous and increasing function of the subunit efficiency indices. The authors prove that “the aggregation requires linearity of the efficiency indices in outputs and inputs jointly”. This result implies that some well-known measures of technical efficiency, such as in Debreu (1951); Farrell (1957) and Färe and Lovell (1978), cannot satisfy the above aggregation condition for any technology set. A quantification of such aggregation inconsistency is provided by Ten Raa (2005). Nevertheless, measures of efficiency can be aggregated by introducing weights (the output shadow prices) as proposed by Li and Ng (1995) and Li and Cheng (2007). The role of weights in the efficiency aggregation is also discussed in Ylvinger (2000). Other measures of aggregate efficiency are described in Bricc et al. (2003) and Färe and Zelenyuk (2003) which generalize to multiple outputs the Farrell (1957) efficiency framework. In the same context, Färe et al. (2004) showed how to aggregate efficiency measures when both the allocative and technical components are considered. Alternate approaches and measures that differ from those of Blackorby and Russell (1999) have been proposed by Cooper et al. (2007). In addition, they have shown how to meet all conditions these authors have specified for satisfactory aggregation of DEA performance evaluations.

5.2 DMUs as components of a greater structure

The models introduced in the previous sections aim at comparing ex-post the efficiency values of different structured DMUs. On the contrary, in this subsection we consider a single DMU that maximizes its future global targets’ achievements by maximizing each subunit contribution to the global objectives. The DMU is usually assumed to pursue its objective by re-allocating resources or fixing targets among its subunits.

5.2.1 Resource allocation

In this framework, a DMU is empowered to reallocate the resources among its subunits such that a global objective is attained instead of possible myopic and specific goals individually achieved by the DMSUs.

Golany and Tamir (1995) proposed a very simple DEA model to allocate m resources among n subunits producing a single output. The objective is to maximize the overall output produced which is equal to the sum of the outputs produced by each subunit. Also, relying on their notation, we denote the i -th input and the output of the DMSU k as \hat{x}_{ik} and \hat{y}_k , respectively, before the reallocation phase takes place. Let B_i be the amount of available resource i , x_{il} and y_l the amount of allocated resource i to DMSU l and the amount of output produced by DMSU l , respectively, once the resource reallocation has been performed.

$$Q^* = \max \sum_l y_l \quad (17a)$$

$$\sum_k \lambda_{kl} \hat{x}_{ik} \leq x_{il} \quad \forall i, l \quad (17b)$$

$$\sum_k \lambda_{kl} \hat{y}_k \geq y_l \quad \forall l \quad (17c)$$

$$\sum_l x_{il} \leq B_i \quad \forall i \quad (17d)$$

$$\sum_k \lambda_{kl} = 1 \quad \forall l \quad (17e)$$

$$\lambda_{kl}, x_{il}, y_l \geq 0 \quad \forall i, k, l. \quad (17f)$$

Constraints (17d) allow to redistribute an amount B_i of resource i among DMSUs so that the production plan of any DMSU l falls inside the production set defined by the observed DMSUs (constraints (17b) and (17c)). The objective (17a), which maximizes the output production, guarantees that the planned DMSUs are on the production frontier.

The theoretical underpinnings of resource allocation models can be found in Bogetoft (1993, 1994). More recently Bogetoft (2000) also introduced an accurate literature analysis and suggested the following principal-agent perspective for governing the relationships among DMUs and their subunits. Asymmetric information exists between a DMU and its subunits: each DMSU knows its production costs exactly, but this information is not granted to the higher level DMU. In this context, a DMU may use DEA-based models to estimate the optimal production possibilities and then limiting the costs that the subunits can claim for their productions. In Bogetoft (2000) regulation problems are considered for a DMU that provides some kind of public utility or service. Agrell et al. (2002) applied the principal-agent perspective in private industry. They suggested an incentive system to promote information sharing as opposite to incentive schemes to regulate costs in the public utilities context.

Many other papers deal with resource reallocation among subunits on the basis of the efficiency values of the different DMUs. Athanassopoulos and Gounaris (2001) considered the Greek hospitals and Athanassopoulos (2004) dealt with the Hellenic Tobacco Organization. Korhonen et al. (2001) provided a description on how to carry out an efficiency analysis of academic research with DEA to support resource reallocation. As an example, they focussed on the data to collect, and suggested how to interpret the results of classical

DEA models. Also for an academic context, Caballero et al. (2004) proposed a three-step algorithm to allocate financial resources. In a first phase, the technical efficiency of each teaching and research unit is evaluated through DEA. On the basis of the efficiency levels obtained, resources are redistributed using a Multi-Criteria Decision Making (MCDM) approach. Finally, the technical efficiency of each DMU is evaluated again after resource redistribution.

Lozano and Villa (2004) proposed models that consider both radial and non-radial measures of efficiency for DMUs and then applied them to the recycling operations (Lozano et al. 2004). In particular, the model that considers a radial measure of efficiency can be seen as the dual of a variable returns to scale version of Model (15) when no additional inputs/output are present. Beasley (2003) considered the problem of allocating both resources and fixed costs. Korhonen and Syrjanen (2004) proposed an iterative procedure to determine a resource allocation plan in the assumption that subunits can slightly modify their production within the current production set.

Ross (2000), dealing with resource allocation in the downstream petroleum industry, proposed a profit maximization model for distribution planning for the DMUs. In this case, the subunits are the delivery vehicles, and their efficiency evaluation is used to set minimal performance requirements on the demand served and on the number of deliveries for each vehicle. Thanassoulis (1996) considered the problem of allocating resources to public services (e.g., schools, hospitals) on the basis of Marginal Resource Levels (MRLs), i.e., “the rates of resource entitlement per unit of each activity or output of an operating unit (e.g., the level of funding per pupil)”. He estimated MRLs as by-products of the efficiency evaluation of the subunits.

Other studies include the evaluation subunits’ efficiencies through classical DEA models in a principal-agent perspective. For example, Chien et al. (2003) dealt with possible reorganizations of power plants. Park and Park (2003) and Ahn and Lee (2004) considered the problem in the supply chain context. Kirkley et al. (2003) showed how the efficiencies of subunits can be used in a regulatory framework that aims at an eco-sustainable development. In particular, they studied the fishery industry.

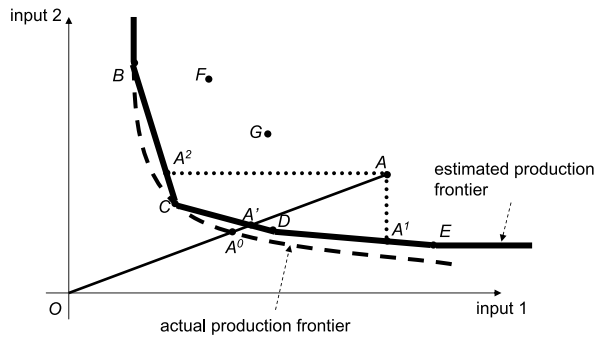
Fang and Zhang (2008) proposed a multiple-objective programming model when the central unit is interested in maximizing both the total efficiency and the efficiency of the subunits by allocating available resources to them.

5.2.2 Target setting

Athanassopoulos (1995) proposed an approach based on targets. In the DEA literature, the targets of a DMU usually are the levels of outputs/inputs that a DMU must reach by increasing/decreasing its current yield/consumption to become efficient. Generally, different targets can be set for an inefficient DMU. In classical DEA models a DMU is required either to reduce all the inputs of the same proportion while keeping fixed the outputs or, vice versa, to increase all the outputs of the same proportion while keeping fixed the inputs. Such choices are strictly related to the definition of a radial measure of the efficiency. As an example, consider Fig. 4.

The classical (constant returns to scale) DEA models would suggest to reduce the inputs of DMU A so that A is projected on A' . Golany (1988) argued that preference information should be considered in setting performance targets. Thanassoulis and Dyson (1992) suggested that the target levels may be fixed differently for each output/input to reflect decision maker preferences. Considering again Fig. 4, DMU A can become efficient by any appropriate reduction of its inputs which projects it on any point of the production frontier

Fig. 4 Projections of DMU A on the actual production frontier and on the estimated production frontier given DMUs A, B, C, D, E, F, G



included between A^1 and A^2 . Athanassopoulos (1995) started from the work by Thanassoulis and Dyson (1992) and developed a model that merges goal programming and data envelopment analysis to compute the optimal targets for each subunit. The model considers two sets of targets. The first set includes targets that are fixed and are possibly not related to the subunit efficiencies but to other objectives, such as efficacy or equity. On the contrary, the second set includes targets concerning the subunit efficiencies. The model assesses the targets by choosing the ones that minimize the deviation of the current performances of the observed DMSUs. Different weights are used to penalize the deviation from the different targets. Athanassopoulos (1995) applied his model to the reallocation of central funds to local authorities of Greece. The same author proposed the use of a DEA model to set targets to power plants (Athanassopoulos et al. 1999). Here, a centralized management uses weights to set different priorities for targets. Similar models are developed in Athanassopoulos (1996) and Athanassopoulos and Triantis (1998). Finally, Athanassopoulos (1998) proposed a four-step procedure that allocates resources according to the principal-agent paradigm. In this context, the subunit efficiencies are evaluated to determine marginal rates of transformation between inputs and outputs of the whole process governed by the DMU. The reader is referred to Athanassopoulos (1998) also for an analysis of the literature on allocation of resources in multilevel public structures.

6 Conclusions

In this work we provide a classification of the main DEA models assessing the efficiency of Decision Making Units when their internal structure is no longer considered as a black box, but insight on their inner processes is available. The interaction in each DMU among the input and output flows and its subunits identifies three broad categories of models. In particular, shared flow models apply when it is possible to partition a DMU as a collection of components that contend their inputs and/or outputs to other components of the same DMU. Multilevel models are referred to when some of the inputs (or outputs) of a DMU are also inputs (or outputs) of its subunits, and some other inputs (or outputs) are not. Furthermore, network models are introduced when intermediate flows among the subunits are taken into account. We show that these formulations are different generalizations of the same elementary model. We extend our analysis to discuss the aggregation of the subunit efficiency indexes into an overall DMU efficiency measure. We finally approach the issue of a single DMU considered as a central authority that reallocates resources or sets predefined targets to its components to maximize its efficiency. Except for the DEA seminal works, Table 1 displays the references that are in common between the papers we analyze. They are organized

according to the sections in which they appear. We observe that our classification is consistent with the references' groupings. Nevertheless, in this work we provide linkages between the different formulations and we encourage the researchers to benefit from previous work for their new speculations.

From a theoretical point of view, the knowledge of the internal structure of DMUs should spot the sources of organizational inefficiency by, e.g., preventing compensations among the subunits. In mathematical terms this translates into linking a DMU and its subunits' efficiencies. This relationship may vary across the different models. But, as a general result, a DMU cannot be efficient if none of its subunits are efficient. Furthermore, several applications show that the discrimination power of a DEA model which considers the internal structure of the DMU always increases with respect to the black box approach. As an extreme case, in some situations all DMUs may turn out to be inefficient.

We conclude the paper by pointing out a few future directions with the understanding that there is large scope of research in the area of this type of DEA models both from a theoretical and application-oriented perspective. Besides the original DEA formulation (Charnes et al. 1978) representing DMUs as black boxes in a constant returns to scale environment, many authors have proposed more sophisticated or alternative approaches taking into account, e.g., nonradial measures of efficiency, value judgements, economic measures of efficiency (see Fried et al. 2008, Chap. 3, for a comprehensive survey of such DEA models). So far, little attention has been devoted to these extensions when DMUs have an identifiable internal structure, especially in the case of multilevel and network models. However, these advances may provide interesting insights, for example the incorporation of intertemporal effects among subunits identifies further sources of inefficiency as shown by Chen (2008) in the case of dynamic network models. Other initial findings in the field of network DEA consider variable returns to scale (Chen et al. 2008) and undesirable factors (Hua and Bian 2008). Then, these new features might also be included into formulations that simultaneously integrate multilevel, network and shared flow characteristics. In this context, Yu and Lin (2008) proposed a first step toward a unified shared flow and network model.

Furthermore, a DMU may pursue its objectives by re-allocating resources or fixing targets among its subunits (see Sect. 5.2). This idea may evolve into two different directions. First, we can consider the possibility that resources are re-allocated and targets are set through DMSU negotiation with each others, and not by decision of the central authority. In this situation, each DMSU maximizes both its efficiency and the one of the whole DMU. Second, resources might be re-allocated and targets set to interdependent DMSUs, as in a supply chain, and not only to homogeneous and independent DMSUs. In this context, the role of asymmetric information between DMSUs and their DMU (see, e.g., Bogetoft 2000) can be extended to the case when such asymmetry exists among DMSUs which make their decisions by means of a negotiation process.

Appendix

Table 1 Synoptic reference table: the papers in the rows are grouped in accordance to the (sub)sections they are cited. Their main references are the papers in the columns

Sections and (subsections)	Färe and Primont (1984)	Yang et al. (2000)	Kao (2000)	Castelli et al. (2004)	Beasley (1995)	Molinero (1996)	Molinero and Tsai (1997)	Färe et al. (1997)	Tsai and Molinero (2002)	Cook et al. (2000)	Cook and Hababou (2001)	Tsai and Molinero (2002)	Diez-Ficio and Mancebon (2002)	Jahanshahloo et al. (2004b)	Jahanshahloo et al. (2004a)	Cook and Green (2004)
1. Elementary models								x								
2. Shared flow models																
Färe and Grosskopf (2000)																
Löthgren and Tambour (1999)																x
Färe and Grosskopf (1996)												x				x
Färe and Whittaker (1995)																
Athanassopoulos (1998)																x
Athanassopoulos (1995)																x
Bogetoft (1994)																
Thanassoulis and Dyson (1992)																
Golany (1988)																
Färe and Zelenyuk (2003)																
Blackorby and Russell (1999)																
Li and Ng (1995)																
Farrell (1957)																x
Debreu (1951)																
Färe and Lovell (1978)																
Cook et al. (1998)																
Cook et al. (2000)																
Molinero and Tsai (1997)																
Molinero (1996)																
Beasley (1995)																
Jahanshahloo et al. (2004b)																x
Jahanshahloo et al. (2004a)																x
Cook and Green (2004)																x

Table 1 (Continued)

Sections and (subsections)					
Färe and Grosskopf (2000)					
Löthgren and Tambour (1999)					
Färe and Grosskopf (1996)					x
Färe and Whittaker (1995)					
Athanassopoulos (1998)					x
Athanassopoulos (1995)			x	x	x
Bogetoft (1994)					x
Thanassoulis and Dyson (1992)	x	x		x	x
Golany (1988)					x
Färe and Zelenyuk (2003)					
Blackorby and Russell (1999)					
Li and Ng (1995)					
Farrell (1957)	x	x	x		
Debreu (1951)					
Färe and Lovell (1978)		x			
Cook et al. (1998)					
Cook et al. (2000)					
Molinero and Tsai (1997)					
Molinero (1996)					
Beasley (1995)					
5.2.2. Target setting					
	Athanassopoulos (1995)				
	Athanassopoulos (1996)				
	Athanassopoulos and Triantis (1998)				
	Athanassopoulos (1998)				
	Athanassopoulos et al. (1999)				

References

- Agrell, P., Bogetoft, P., & Tind, J. (2002). Incentive plans for productive efficiency, innovation and learning. *International Journal of Production Economics*, 78(1), 1–11.
- Ahn, H., & Lee, H. (2004). An agent-based dynamic information network for supply chain management. *BT Technology Journal*, 22(2), 18–27.
- Amirteimoori, A., & Nashtaei, R. (2006). The role of time in multi-component efficiency analysis: An application. *Applied Mathematics and Computation*, 177(1), 11–17.
- Amirteimoori, A., & Shafiei, M. (2006). Measuring the efficiency of interdependent decision making sub-units in DEA. *Applied Mathematics and Computation*, 173(2), 847–855.
- Athanassopoulos, A. (1995). Goal programming & data envelopment analysis (Go DEA) for target-based multi-level planning: allocating central grants to the Greek local authorities. *European Journal of Operational Research*, 87(3), 535–550.
- Athanassopoulos, A. (1996). Assessing the comparative spatial disadvantage (csd) of regions in the European Union using non-radial data envelopment analysis methods. *European Journal of Operational Research*, 94(3), 439–452.
- Athanassopoulos, A. (1998). Decision support for target-based resource allocation of public services in multi-unit and multilevel systems. *Management Science*, 44(2), 173–187.
- Athanassopoulos, A. (2004). Using frontier efficiency models as a tool to re-engineer networks of public sector branches: An application to the Hellenic Tobacco Organization. *European Journal of Operational Research*, 154(2), 533–547.
- Athanassopoulos, A., & Gounaris, C. (2001). Assessing the technical and allocative efficiency of hospital operations in Greece and its resource allocation implications. *European Journal of Operational Research*, 133(2), 416–431.
- Athanassopoulos, A., Lambroukos, N., & Seiford, L. (1999). Data envelopment scenario analysis for setting targets to electricity generating plants. *European Journal of Operational Research*, 115(3), 413–428.
- Athanassopoulos, A., & Triantis, K. (1998). Assessing aggregate cost efficiency and the related policy implications for Greek local municipalities. *INFOR*, 36(3), 66–83.
- Banker, R., Charnes, A., & Cooper, W. (1984). Some models for estimating technical and scale inefficiencies in DEA. *Management Science*, 30(9), 1078–1092.
- Beasley, J. (1995). Determining teaching and research efficiencies. *Journal of the Operational Research Society*, 46(4), 441–452.
- Beasley, J. (2003). Allocating fixed costs and resources via data envelopment analysis. *European Journal of Operational Research*, 147(1), 198–216.
- Blackorby, C., & Russell, R. (1999). Aggregation of efficiency indices. *Journal of Productivity Analysis*, 12(1), 5–20.
- Bogetoft, P. (1993). Parallelism in information production: Moral hazard and relative performance evaluations. *Management Science*, 39(4), 448–457.
- Bogetoft, P. (1994). Incentive efficient production frontiers: An agency perspective on DEA. *Management Science*, 40(8), 959–968.
- Bogetoft, P. (2000). DEA and activity planning under asymmetric information. *Journal of Productivity Analysis*, 13(1), 7–48.
- Briec, W., Dervaux, B., & Leleu, H. (2003). Aggregation of directional distance functions and industrial efficiency. *Journal of Economics—Zeitschrift für Nationalökonomie*, 79(3), 237–261.
- Caballero, R., Galache, T., Gomez, T., Molina, J., & Torrico, A. (2004). Budgetary allocations and efficiency in the human resources policy of a university following multiple criteria. *Economics of Education Review*, 23(1), 67–74.
- Castelli, L., Pesenti, R., & Ukovich, W. (2001). DEA-like models for efficiency evaluations of specialized and interdependent units. *European Journal of Operational Research*, 132(2), 274–286.
- Castelli, L., Pesenti, R., & Ukovich, W. (2004). DEA-like models for the efficiency evaluation of hierarchically structured units. *European Journal of Operational Research*, 154(2), 465–476.
- Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
- Charnes, A., Cooper, W., Golany, B., Seiford, L., & Stutz, J. (1985). Foundation of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30(12), 91–107.
- Chen, C.-M. (2008, in press). A network-DEA model with new efficiency measures to incorporate the dynamic effect in production networks. *European Journal of Operational Research*. doi:10.1016/j.ejor.2007.12.025.
- Chen, Y., Cook, W., Ning, L., & Zhu, J. (2008, in press). Additive efficiency decomposition in two-stage DEA. *European Journal of Operational Research*. doi:10.1016/j.ejor.2008.05.011.

- Chen, Y., Liang, L., & Yang, F. (2006). A DEA game model approach to supply chain efficiency. *Annals of Operations Research*, 145(1), 5–13.
- Chen, Y., Liang, L., & Zhu, J. (2007, in press). Equivalence in two-stage DEA approaches, *European Journal of Operational Research*. doi:10.1016/j.ejor.2007.11.040.
- Chen, Y., & Zhu, J. (2004). Measuring information technology's indirect impact on firm performance. *Information Technology and Management*, 5(12), 9–22.
- Chien, C., Lo, F., & Lin, J. (2003). Using DEA to measure the relative efficiency of the service center and improve operation efficiency through reorganization. *IEEE Transactions on Power Systems*, 18(1), 366–373.
- Cook, W., Chai, D., Doyle, J., & Green, R. (1998). Hierarchies and groups in DEA. *Journal of Productivity Analysis*, 10(2), 177–198.
- Cook, W., & Green, R. (2004). Multicomponent efficiency measurement and core business identification in multiplant firms: A DEA model. *European Journal of Operational Research*, 157(3), 540–551.
- Cook, W., & Green, R. (2005). Evaluating power plant efficiency: A hierarchical model. *Computers and Operations Research*, 4(9), 813–823.
- Cook, W., & Hababou, M. (2001). Sales performance measurement in bank branches. *Omega-International Journal of Management Science*, 29(4), 299–307.
- Cook, W., Hababou, M., & Tuenter, H. (2000). Multicomponent efficiency measurement and shared inputs in data envelopment analysis: An application to sales and service performance in bank branches. *Journal of Productivity Analysis*, 14(3), 209–224.
- Cook, W., & Zhu, J. (2005). *Modeling performance measurement: applications and implementation issues in DEA*. Berlin: Springer.
- Cooper, W., Huang, Z., Li, S., Parker, B., & Pastor, J. (2007). Efficiency aggregation with enhanced Russell measures in data envelopment analysis. *Socio-Economic Planning Sciences*, 41(1), 1–21.
- Cooper, W., Seiford, L., & Tone, K. (2000). *Data envelopment analysis*. Boston: Kluwer Academic.
- Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19(3), 273–292.
- Diez-Ticio, A., & Mancebon, M. (2002). The efficiency of the Spanish police service: an application of the multiactivity DEA model. *Applied Economics*, 34(3), 351–362.
- Fang, L., & Zhang, C.-Q. (2008). Resource allocation based on the DEA model. *Journal of the Operational Research Society*, 59(8), 1136–1141.
- Färe, R., Grabowski, R., Grosskopf, S., & Kraft, S. (1997). Efficiency of a fixed but allocable input: A non-parametric approach. *Economics Letters*, 56(2), 187–193.
- Färe, R., & Grosskopf, S. (1996). Productivity and intermediate products: A frontier approach. *Economics Letters*, 50(1), 65–70.
- Färe, R., & Grosskopf, S. (2000). Network DEA. *Socio-Economic Planning Sciences*, 34(1), 35–49.
- Färe, R., Grosskopf, S., & Whittaker, G. (2007). Network DEA. In *Modeling data irregularities and structural complexities in DEA* (pp. 209–240). New York: Springer.
- Färe, R., Grosskopf, S., & Zelenyuk, V. (2004). Aggregation of cost efficiency: indicators and indexes across firms. *Academia Economic Papers*, 32(3), 395–411.
- Färe, R., & Lovell, C. (1978). Measuring the technical efficiency of production. *Journal of Economic Theory*, 19(1), 150–162.
- Färe, R., & Primont, D. (1984). Efficiency measures for multiplant firms. *Operations Research Letters*, 3(5), 257–260.
- Färe, R., & Whittaker, G. (1995). An intermediate input model of dairy production using complex survey data. *Journal of Agricultural Economics*, 46(2), 201–213.
- Färe, R., & Zelenyuk, V. (2003). On aggregate Farrell efficiencies. *European Journal of Operational Research*, 146(3), 615–620.
- Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A, General*, 120(3), 253–281.
- Fried, H., Knox Lovell, C., & Schmidt, S. (2008). *The measurement of productive efficiency and productive growth*. Oxford: Oxford University Press.
- Golany, B. (1988). An interactive MOLP procedure for the extension of DEA to effectiveness analysis. *Journal of the Operational Research Society*, 39(8), 725–734.
- Golany, B., Hackman, S., & Passy, U. (2006). An efficiency measurement framework for multi-stage production systems. *Annals of Operations Research*, 145(1), 51–68.
- Golany, B., & Tamir, E. (1995). Evaluating efficiency-effectiveness-equality trade-offs: A data envelopment approach. *Management Science*, 41(7), 1172–1184.
- Hua, Z., & Bian, Y. (2008). Performance measurement for network dea with undesirable factors. *International Journal of Management and Decision Making*, 9(2), 141–153.
- Jahanshahloo, G., Amirteimoori, A., & Kordrostami, S. (2004a). Measuring the multi-component efficiency with shared inputs and outputs in data envelopment analysis. *Applied Mathematics and Computation*, 155(1), 283–293.

- Jahanshahloo, G., Amirteimoori, A., & Kordrostami, S. (2004b). Multi-component performance, progress and regress measurement and shared inputs and outputs in DEA for panel data: an application in commercial bank branches. *Applied Mathematics and Computation*, *151*(1), 1–16.
- Kao, C. (2000). Short-run and long-run efficiency measures for multiplant firms. *Annals of Operations Research*, *97*(1), 379–388.
- Kao, C. (2007, in press). Efficiency decomposition in network data envelopment analysis: A relational model. *European Journal of Operational Research*. doi:10.1016/j.ejor.2007.10.008.
- Kao, C. (2008, in press). Efficiency measurement for parallel production systems. *European Journal of Operational Research*. doi:10.1016/j.ejor.2008.04.020.
- Kao, C., & Hwang, S. (2008). Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *European Journal of Operational Research*, *185*(1), 418–419.
- Kirkley, J., Squires, D., Alam, M., & Ishak, H. (2003). Excess capacity and asymmetric information in developing country fisheries: The Malaysian purse seine fishery. *American Journal of Agricultural Economics*, *85*(3), 647–662.
- Korhonen, P., & Syrjanen, M. (2004). Resource allocation based on efficiency analysis. *Management Science*, *50*(8), 1134–1144.
- Korhonen, P., Tainio, R., & Wallenius, J. (2001). Value efficiency analysis of academic research. *European Journal of Operational Research*, *130*(1), 121–132.
- Lewis, H., & Sexton, T. (2004). Network DEA: efficiency analysis of organizations with complex internal structure. *Computers and Operations Research*, *31*(9), 1365–1410.
- Li, S., & Cheng, Y. (2007). Solving the puzzles of structural efficiency. *European Journal of Operational Research*, *180*(2), 713–722.
- Li, S., & Ng, Y. (1995). Measuring the productive efficiency of a group of firms. *International Advances in Economic Research*, *1*(4), 377–390.
- Liang, L., Yang, F., Cook, W., & Zhu, J. (2006). Dea models for supply chain efficiency evaluation. *Annals of Operations Research*, *145*(1), 35–49.
- Löthgren, M., & Tambour, M. (1999). Productivity and customer satisfaction in Swedish pharmacies. A DEA network model. *European Journal of Operational Research*, *115*(3), 449–458.
- Lozano, S., & Villa, G. (2004). Centralized resource allocation using data envelopment analysis. *Journal of Productivity Analysis*, *22*(1), 143–161.
- Lozano, S., Villa, G., & Adenso-Diaz, B. (2004). Centralised target setting for regional recycling operations using DEA. *Omega—International Journal of Management Science*, *32*(2), 101–110.
- Molinero, C. (1996). On the joint determination of efficiencies in a data envelopment analysis context. *Journal of the Operations Research Society*, *47*(10), 1273–1279.
- Molinero, C., & Tsai, P. (1997). Some mathematical properties of a DEA model for the joint determination of efficiencies. *Journal of the Operations Research Society*, *48*(1), 51–56.
- Park, J., & Park, S. (2003). Agent-based merchandise management in business-to-business electronic commerce. *Decision Support Systems*, *35*(3), 311–333.
- Prieto, A., & Zofio, J. (2007). Network DEA efficiency in input-output models: With an application to OECD countries. *European Journal of Operational Research*, *178*(1), 292–304.
- Ross, A. (2000). Performance-based strategic resource allocation in supply networks. *International Journal of Production Economics*, *63*(3), 255–266.
- Seiford, L., & Zhu, J. (1999). Profitability and marketability of the top 55 u.s. commercial banks. *Management Science*, *45*(9), 1270–1288.
- Sexton, T., & Lewis, H. (2003). Two-stage DEA: An application to major league baseball. *Journal of Productivity Analysis*, *19*(2), 227–249.
- Shephard, R. (1970). *Theory of cost and production frontier*. Princeton: Princeton University Press.
- Simar, L., & Wilson, P. (2000). Statistical inference in non parametric frontier models: the state of the art. *Journal of Productivity Analysis*, *13*(1), 49–78.
- Staat, M. (2001). The effect of sample size on the mean efficiency in DEA: Comment. *Journal of Productivity Analysis*, *15*(2), 129–137.
- Staat, M. (2002). Bootstrapped efficiency estimates for a model for groups and hierarchies in dea. *European Journal of Operational Research*, *138*(1), 1–8.
- Ten Raa, T. (2005). Aggregation of productivity indices: The allocative efficiency correction. *Journal of Productivity Analysis*, *24*(2), 203–209.
- Thanassoulis, E. (1996). A data envelopment analysis approach to clustering operating units for resource allocation purposes. *Omega—International Journal of Management Science*, *24*(4), 463–476.
- Thanassoulis, E., & Dyson, R. (1992). Estimating preferred target input-output levels using data envelopment analysis. *European Journal of Operational Research*, *56*(1), 80–97.

- Thompson, R., Langemeier, L., Lee, C., Lee, E., & Thrall, R. (1990). The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Econometrics*, *46*(12), 93–108.
- Tsai, P., & Molinero, C. (2002). A variable returns to scale data envelopment analysis model for the joint determination of efficiencies with an example of the UK health service. *European Journal of Operational Research*, *141*(1), 21–38.
- Tulkens, H., & Eeckaut, P. (1995). Non-parametric efficiency progress and regress measures for panel data: methodological aspects. *European Journal of Operational Research*, *80*(3), 474–499.
- Yang, Y., Ma, B., & Koike, M. (2000). Efficiency-measuring DEA model for production system with k independent subsystems. *Journal of the Operations Research Society of Japan*, *43*(3), 343–354.
- Ylvinger, S. (2000). Industry performance and structural efficiency measures: Solutions to problems in firm models. *European Journal of Operational Research*, *121*(1), 164–174.
- Yu, M.-M., & Lin, E. (2008). Efficiency and effectiveness in railway performance using a multi-activity network DEA model. *Omega—International Journal of Management Science*, *36*, 1005–1017.
- Zhang, Y., & Bartles, R. (1998). The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand. *Journal of Productivity Analysis*, *9*(2), 187–204.